How to Make Experimental Economics Research More Reproducible:

Lessons from Other Disciplines and a New Proposal[1]

Zacharias Maniadis, University of Southampton

Fabio Tufano, University of Nottingham

John A. List University of Chicago[2]

Abstract: Efforts in the spirit of this special issue aim at improving the reproducibility of experimental economics, in response to the recent discussions regarding the "research reproducibility crisis." We put this endeavour in perspective by summarizing the main ways (to our knowledge) that have been proposed – by researchers from several disciplines – to alleviate the problem. We discuss the scope for economic theory to contribute to evaluating the proposals. We argue that a potential key impediment to replication is the expectation of negative reactions by the authors of the individual study, and suggest that incentives for having one's work replicated should increase.

JEL Codes: B40, C90

Keywords: False-Positives, Reproducibility, Replication

Introduction

Nearly a century ago Ronald Fisher began laying the groundwork of experimental practice, which remains today. Fisher's work was highlighted by the experimental tripod: the concepts of replication, blocking, and randomization were the foundation on which the analysis of the experiment was based. The current volume of *Research in Experimental Economics* (REE for short) aims at helping experimental economists think more deeply about replication, which has attracted the attention of scholars in other disciplines for some time. In particular, in many biological and human sciences it has been claimed there is a 'reproducibility crisis', whereby established results fail to be replicated (Jennions and Moller, 2002; Ioannidis, 2005; Nosek, Spies, and Motyl, 2012; Bettis, 2012).

[1]This chapter has been published in C.A. Deck, E. Fatas, & T. Rosenblat (Eds.), Replication in Experimental Economics (Vol. 18) , Research in Experimental Economics, Bingley, UK, Emerald Group Publishing Limited.
[2]Maniadis: Dept. of Economics, School of Social Sciences, University of Southampton, Southampton SO17 1BJ, UK (email: Z.Maniadis@soton.ac.uk). Tufano: CeDEx, School of Economics, University of Nottingham, University Park, Nottingham NG7 2RD, UK (e-mail: fabio.tufano@nottingham.ac.uk). List: Dept. of Economics, University of Chicago, 1126 E.59th Street, Chicago, IL 60637, USA (e-mail: jlist@uchicago.edu).

Even the popular press has joined the discussion, which has in turn attracted the interest of the general public.[3] In contrast, the discussion regarding the importance of replication in experimental economics has only recently been gaining traction (e.g., Maniadis, Tufano and List; 2014, Coffman and Niederle, 2015). Another initiative in experimental economics is the launch of the new *Journal of the Economic Science Association*, which strives to promote replication studies.

This article aims to place the endeavour of the current REE volume in context. When it comes to the big picture, the main question is what types of reforms may improve the credibility and replicability of experimental findings. Initiatives like this volume are one of several possible ways in which a culture of frequent replications can be promoted. First, we wish to provide the economics audience with an overview of the solutions and initiatives for the larger problem that have been proposed in comparable disciplines.[4] Second, we discuss the challenges in rigorously evaluating these proposals. Third, we propose a novel solution for inducing replications that uniquely tackles an important issue that is neglected in other approaches to reform: the disincentives of original authors to cooperate with scholars who wish to replicate their studies.

We envision that this analysis will has certain benefits. First, independent scholars as well as interested parties who wish to tackle a possible credibility problem in economics – many of the readers of this volume – will be exposed to several important ideas that are already out there but are difficult to access across disciplinary boundaries. Second, there is a great need for economic theorists to formalize these proposals. As respectable experts within their own 'ivory towers', scholars in each discipline are inclined to propose informally solutions that seem promising. At the same time, there is a scarcity of formal and systematic evaluation of those solutions. As John Ioannidis, a leading proponent of meta-research on the replicability problem notes: "... it is essential that we obtain as much rigorous evidence as possible, including experimental studies, on how these practices perform in real life and whether they match their theoretical benefits [...]. Otherwise, we run the risk that we may end up with worse scientific credibility than in the current system" (Ioannidis, 2012). Finally, we believe that a crucial impediment for a "healthy" culture of frequent replications, namely the disincentive to have one's work replicated, has not been addressed by the current proposals and our suggestion might contribute to a much needed discussion of incentives. We shall first recite the existing proposals for reform and the challenges to evaluate them; then, we describe our modest proposal.

Four Categories of Proposals

Proposed solutions can be placed into several different categories. Some of them focus on improving the accuracy of inference that can be made on the basis of received empirical

[3] For instance, there have been articles in The Guardian (Goldacre, 2011), The Economist (2013, 2014), The New York Times (Zimmer, 2011) and The New Yorker (Lehrer, 2010).

[4] In doing so, without claiming exhaustiveness, we intend to also highlight to the interested reader additional sources reporting on relevant scholarly debates and reform proposals advanced. For example, two recent special issues of the psychology journal *Perspectives on Psychological Science* have been devoted to replicability (vol. 7, issue 6, November 2012; and vol. 9, issue 3, May 2014).

results. Some arguments and initiatives wish to tie researchers' hands in order to prevent excess flexibility: they propose to make pre-registration of studies obligatory. Others mainly concern reforms to raise disclosure standards in reporting empirical research. Finally, there are proposals for increasing the number of replications. After describing the different suggestions, we discuss efforts and difficulties in evaluating their effects in a rigorous scientific framework.

Pre-Registering Studies

A prominent proposed solution is to force researchers to submit a full plan of their design and analysis prior to conducting the experiment.[5] Miguel et al. (2014) note that although clinical researchers in the United States are forced by law since 2007 to submit their studies in a public database and to post a review of the results, only recently have social scientists followed this example. This is a situation where the outcome of the policy for social welfare is unclear. There is a cost in satisfying the requirements of pre-registration, and many people argue that it is likely to stifle creativity and harm exploratory studies. Others argue that what will happen is simply that people will clearly designate exploratory study for what it is, rather than presenting it as hypothesis-driven research (Miguel et al., 2014).

Important issues shall be able to be clarified by means of formal theory. For instance, Henry (2009) introduces a model of strategic disclosure of research results. Interestingly, he finds that mandatory disclosure rules, such as pre-registration, may be either welfare enhancing or reducing depending on specific cases in which information is either withheld or shared. Similarly, Coffman and Niederle (2014) posit that the benefits of pre-registration largely depend on the specificities of the experimental setting: indeed, they suggest that pre-registration will only be fruitful in extreme cases, such as large and expensive field experiments. To generate such theoretical insights, they use the framework by Ioannidis (2005). Their main idea is that pre-registration operates only through reducing the generic research "bias" of each study, but not the selection of positive findings to be reported after multiple testing. Yet, it is clear that theory formulation significantly enhances our capacity of engaging in careful debate. For example, regarding Coffman and Niederle (2014), one could argue that if pre-registration is connected with funding, or otherwise generates a reputation cost in leaving the study unpublished then the model with 'research competition' is no longer appropriate. The reason is that as long as there is at least one pre-registration, the results of the given study will necessarily be published, and it is no longer true that only a positive result will be published (as assumed in the version of the model with k competitive research teams).

[5] In a similar spirit, Dufwenberg and Martinsson (2014) argue in favour or changing the submission process. They maintain that it is only the design that should be submitted, but not the results. This is likely to have an equivalent outcome as pre-registration in the sense that it will disconnect the incentive to publish with how the data look like. It also seems to strike both problems (biased methodological choices at the researcher level and publication bias at the macro level) simultaneously.

Raising Disclosure Standards

A large set of authors in both the social and the behavioural sciences has argued that the best way to increase the credibility of experimental results is to force researchers to reveal all relevant information that is required in order to make appropriate inference from the experimental studies. In particular, it has been suggested that each experiment must be accompanied by a complete list of important informational pieces that the researcher needs to provide (Simmons, Nelson and Simonsohn, 2011; Landis et al., 2012; Fanelli, 2013, Miguel et al., 2014). If researchers truthfully complete this list, then appropriate inference from the generated data can be substantially improved.

Many authors have considered this idea. For example, Miguel et al. (2014) note that disclosure standards "… require researchers to report all measures, manipulations and data exclusions, as well as how they arrived at final sample sizes." Simmons, Nelson and Simonsohn (2011) illustrate thoroughly how easy it is to experimentally prove the existence of a false effect, if essential information is absent. Landis et al. (2012) report evidence that many preclinical experiments are poorly reported, and that this is correlated with overstated findings. They also note that many involved parties and stakeholders need to work together in order to achieve increased transparency in preclinical research. Fanelli (2013) proposes to redefine misconduct in such a way that it incorporates mild forms of questionable research practices (QRP), hence promoting transparency.[6]

The idea behind disclosure requirements is that people will honestly follow them. In particular, it has been argued that the enforcement of the disclosure list will transform 'sins of omission' into 'sins of commission', and hence they will be avoided (Simmons, Nelson and Simonsohn, 2011). In other words, it might be true that researchers may exploit the loopholes of the current system by engaging in QRP, but it is also reasonable to expect that otherwise honest people are not likely to engage in pure deception after the regime changes.[7] Despite the plausibility of this argument, it is worth illustrating the difficulty in evaluating this proposal. Given that the set of incentives will change after the reform, it is difficult to predict how things will evolve. Striking down mild transgressions might make people turn to hard forms of misbehaviour, and the overall effect remains unclear.

In an example of how economic theory can help, Gall and Maniadis (2014) show that enforcing a transparency requirement may reduce the overall rate of misconduct. The authors examine an environment with limited attention and a restricted number of top journals where researchers are intrinsically motivated to adhere to honesty in performing research, but are also interested in being published well. There are three types of behaviour for researchers who wish to publish their results: presenting the results as they are, engaging in 'mild' QRP,[8]

[6] A radical approach with respect to transparency comes from new 'workflow tools' that allow researchers to analyse their data, while making their results and workflow accessible to others. Poldrack and Poline (2014) discuss how this idea has been applied in genomics research.

[7] This notion is also supported by the survey evidence in Fanelli (2009), who shows that pure scientific fraud is much less common relative to mild forms of questionable research practices.

[8] A set of survey studies identifies types of questionable research practices that researchers are likely to engage in (List et. al., 2001; Fanelli, 2009; John, Loewenstein and Prelec, 2012). Nosek, Spies and Motyl (2012)

and engaging in outright deception. The publication game is modelled as a 'rat race', in the sense that the more others cheat, the less likely a given researcher is to publish her work. The authors find that somewhat counterintuitively, ruling out mild misbehaviour (which is the outcome of enforcing disclosure) robustly decreases the overall rate of misconduct.[9]

We believe that this type of theory is necessary in order to structure the arguments and especially for considering general equilibrium effects of policy changes. There is clearly an open and promising area for research here. The challenge is to have the theory guide our empirical research of the likely effect of the regime change to start a virtuous circle between theoretical and empirical investigations in order to derive possible policy implications.

Sharing Data and Materials

It has been argued that sharing data and materials would increase transparency in science (e.g., Miguel et al., 2014); however, the implementation of such an apparently straightforward change has been at times quite complex and has generated a heated trans-disciplinary debate. For instance, legitimate questions can be posed about who should set the sharing rules; who should be expected to enact such rules; what type of data/materials should be shared (e.g., raw versus processed data); who should have access to the data; how third-party access should be granted, and the like (Doshi, Goodman and Ioannidis, 2013).

In recent years, significant changes have been put in place in economics. For instance, the data availability policies of top journals, such as *The American Economic Review* and *The Review of Economic Studies* currently require that, prior to publication authors of accepted manuscripts must provide data and other relevant materials in order to make them publically available on the journal's repositories. However, such policies are not yet in place for most other scientific outlets in economics.

Along with journals' repositories, new archiving tools have been developed by several stakeholders to facilitate the sharing of data and materials. This is, for instance, the case with the 'Open Science Framework', an online initiative by the 'Center for Open Science' aimed at archiving data and materials.[10] Moreover, even when those data and materials are available, reproducibility of published results (also called "routine replication" – see Maniadis, Tufano and List, 2014) may be not easily obtained. In genetics, Ioannidis et al. (2009) demonstrate that despite the public availability of data, the reproducibility of published results by independent scholars – who are experts of the relevant fields – is limited.

In economics, Dewald, Thursby, and Anderson (1986) report on a data availability project implemented by the *Journal of Money, Credit and Baking*: their main finding is that the

---

forcefully argue that human cognition works in such ways that it is entirely ordinary behaviour that may distort the scientific evidence, for example by post-hoc hypothesizing and eventually forgetting that the results were unexpected.

[9] Of course, the results of Gall and Maniadis (2014) hinge on a few key assumptions, notably the 'tournament structure' of the publication game and the higher cost for hard forms of QRP relative to soft forms of QRP. Clearly, it is also important to evaluate the effects of imposing the disclosure requirement under different modelling environments.

[10] For details, see https://osf.io/ [Accessed on 13 April 2015].

frequent occurrence of inadvertent errors makes reproducibility close to impossible. Thus, it seems that sharing of data and materials may surely improve transparency, and perhaps attenuate critical obstacles to replication, but it is not the panacea to all the reproducibility problems of scientific research.

Other Ideas

The three large categories of proposals recited above are not exclusive. For example, Ioannidis (2014) provides a variety of ideas for reform that warrant careful thought and perhaps formalization. An intriguing one is to reduce the return from obtaining grants and occupying positions of power in terms of academic recognition. On the contrary, he envisions a regime where these resources are 'penalized', viewed as opportunities that the researcher will have to utilize by increased productivity. On a final positive note, Nosek, Spies and Motyl (2012) are proponents of 'paradigmatic-driven' research, whereby the research material is standardized and altered in a very structured way for each new piece of research. It seems to us that experimental economics already achieves this standard relative to other disciplines such as psychology (Hertwig and Ortmann, 2001).

Encouraging a Replication Culture

The practices of raising disclosure standards and of pre-registration target research bias, which affects reported results. The third possibility is to trust the traditional method for ensuring the truthfulness of research findings: replication. Replication is perhaps the most powerful tool that we have to change priors about a behavioural relationship between two variables. To see this point, consider the framework in Maniadis, Tufano and List (2014), where n represents the number of associations that are being studied in a specific field, π is the fraction of these associations that are actually true, α denotes the typical significance level, and $1-\beta$ denotes the typical power of the experimental design. Scholars are typically interested in the Post-Study Probability (PSP) that the research finding is true.

The PSP can be found as follows: of the n examined associations, πn associations will be true, and $(1-\pi)n$ will be false. Among the true ones, $(1-\beta)\pi n$ will be declared as true, whereas among the false associations, $\alpha(1-\pi)n$ will be declared true even though they are false (false positives). The PSP is simply found by dividing the number of true associations which are declared true by the number of all associations declared true:

[1] $$PSP = \frac{(1-\beta)\pi}{(1-\beta)\pi+\alpha(1-\pi)}$$

Maniadis, Tufano and List (2014) use this framework to show the power of replication.11 To illustrate, we consider their Table 5, reproduced here as Table 1. The authors calculate the probability that anywhere from zero (i = 0 in Table 1) to four investigations (i = 3 in Table 1) reject the null, given that the relationship is true and given that it is false. They derive the

[11] Such Bayesian arguments have a long history in the biomedical sciences. Similar points in the context of these sciences were made in Wacholder et al., 2004; Ioannidis, 2005; Moonesinghe, Khoury, and Janssens, 2007.

PSP as the fraction of the true associations over all associations for each level of replication (bold numbers indicate cases where the PSP is less than half). Empirical results reported in Table 1 show that with just a few independent positive replications, the improvement in PSP is considerable, especially in those cases where the original result was a 'surprise' or unexpected.

Exact replications not only show us if the initial results are a by-product of QRP, but also ensure that the inherent randomness and context-specificity of human behaviour were not the driver of an initial finding. Yet, it is well known that the incentives for performing replications are weak (Ioannidis, 2005; Nosek, Spies and Motyl, 2012).

In one of the few models examining incentives for replication, Kiri, Lacetera and Zirulia (2014) employ a game-theoretic model to demonstrate that a fraction of non-reproducible studies has to be expected in any state of the world. In particular, if in a given field no such studies are identified, their model implies that there must be insufficient replication efforts. They show formally that incentives of would-be replicators play a key role in fostering replication. On the other hand, according to the model, the frequently raised calls for softening the overall incentives to publish do not manage to affect research quality as hypothesized. Much less is known, however, about the exact mechanisms and incentive types for would-be replicators and authors of the initial study.

	Power = 0.80				Power = 0.50				Power = 0.20			
π	i=0	i=1	i=2	i=3	i=0	i=1	i=2	i=3	i=0	i=1	i=2	i=3
	PSP											
0.01	**0.02**	**0.1**	**0.47**	0.91	**0.02**	**0.1**	**0.45**	0.89	**0.02**	**0.07**	**0.22**	0.54
0.02	**0.05**	**0.19**	0.64	0.95	**0.05**	**0.19**	0.63	0.94	**0.04**	**0.13**	**0.36**	0.71
0.05	**0.12**	**0.38**	0.82	0.98	**0.12**	**0.38**	0.81	0.98	**0.1**	**0.28**	0.6	0.86
0.1	**0.22**	0.56	0.91	0.99	**0.22**	0.56	0.9	0.99	**0.2**	**0.45**	0.76	0.93
0.2	**0.38**	0.74	0.96	1	**0.38**	0.74	0.95	1	**0.36**	0.64	0.88	0.97
0.35	0.57	0.86	0.98	1	0.57	0.86	0.98	1	0.55	0.8	0.94	0.98
0.55	0.75	0.93	0.99	1	0.75	0.93	0.99	1	0.73	0.9	0.97	0.99

Table 1—The PSP Estimates as a Function of Prior Probability (π), Power and Number of Replications (i)

Despite the relative lack of evidence, there have been numerous proposals and approaches. The first idea with respect to how replication can be enhanced is embodied in this special REE volume and also in the new *Journal of Economic Science Association*: increase the number of outlets that are friendly to replication studies. Nosek, Spies and Motyl (2012), however, argue that this model may not have strong promise, since if replication continues to be considered a low-status endeavour, then these outlets may themselves be characterized as such and researchers might have low interest in publishing in such journals.

An alternative idea is conducting structured sets of replications on the basis of pre-registration (Nosek and Lakens, 2014) which may incentivise replications by guaranteeing

publication.[12] Moreover, in psychology large replication projects with the participation of many collaborators have been organized (called "the Many Labs' Replication Project", Klein et al., 2014). This provides a paradigm that could also be used for replicating well-known experimental economics results.[13] In fact, in this vein, the "Behavioural Economics Replication Project" is currently undertaking the replication of eighteen prominently published studies in the field of 'psychology and economics'.[14] It has also been proposed that the editors themselves should make direct replication a necessary prerequisite for publication (Nosek, Spies and Motyl, 2012; Begley and Ellis, 2012). Nosek, Spies and Motyl (2012) note that a direct replication might be much easier and practical to enforce if the results are very surprising and the data collection is relatively cheap.

A New Proposal: Addressing the Incentives of Having One's Work Replicated

There are still many barriers to a healthy replication culture. Some are institutional; others relate to the incentive system, and many others are cultural. When it comes to replication the perception is often that a "witch hunt" is taking place. Initial authors, whose studies are under scrutiny and subject to a replication attempt, feel under attack.[15] This is the case also because – both within the scientific community and beyond it – a failed replication often is interpreted as implying either a poorly conducted study or even malfeasance and scientific misconduct from the part of the researchers of the initial study.

However, this should not be the case. As is clear from equation (1) above, failed replications should be viewed as part of the mechanics of empirical inference, which is bound to yield a certain number of false positives (e.g., Ioannidis, 2005; Maniadis Tufano and List, 2014; Coffman and Niederle, 2015). On the other hand, replicators' attitude and behaviour often may be interpreted as particularly threatening.[16] Thus, original authors face significant costs and little expected returns from their active collaboration with would-be replicators. The dynamics between original authors and would-be replicators – with no surprise – may be at

[12] Nosek And Lakens (2014) introduce the novel endeavour as follows: "This article introduces the first known journal issue in any discipline consisting exclusively of preregistered replication studies. It demonstrates that replications have substantial value, and that incentives can be changed." They note that the advantage of pre-registration lies in guaranteeing publication, and also in addressing a subtle bias of the referees called CARKing: **"**When the results are known, evaluation of quality is likely influenced by preexisting beliefs. […] Motivated reasoning makes it easy to generate stories for why results differed from expectations.**"** Pre-registration of the replications addresses this concern by basing the review on the design only.

[13]A similar proposal has been made in the field of neuroscience by Wagenmakers and Forstmann (2014), who argue that editors should have an active role in soliciting replication attempts of famous studies with full pre-registration.

[14] The project involves scholars from several institutions (i.e., Caltech, the University of Innsbruck, the Stockholm School of Economics and the National University of Singapore). For further details, see http://sciencepredictionmarkets.com/ [last accessed on 10 April 2015].

[15] It seems that there is a lack of maturity in both scientists and the general public about how to respond to replication. Poldrack and Poline (2014) note: "First, it is important to clearly distinguish the lack of reproducibility (even if due to errors) from scientific misconduct. There has been an unfortunate tendency (especially amongst online discussions) to conflate these two issues, which can lead to a toxic atmosphere of mistrust and accusation that undermines any effective scientific discourse."

[16] An 'etiquette for replication' has been put forward (e.g., Kahneman, 2014) in order to curb this alleged "bullying" by replicators.

times tough and irksome.[17] This may constitute a further disincentive to promoting a culture of frequent well-conducted replications, or even to cultivate replication as a "social norm," which is followed with the confidence that (most) members of our scientific community will do the same (e.g., Bicchieri, 2006). To foster such a shift and make replication a healthy "social norm" in our scientific community we expect that several measures should be put in place, after positive appropriate evaluations. Several scholars have focused their attention on would-be replicators' incentives (e.g. Nosek, Spies and Motyl, 2012) to promote replications and knowledge accumulation. However, this approach may prove insufficient, or simply result in an excessively slow evolution towards a norm of frequent replication (e.g., Cavalli-Sforza and Feldman, 1981).

We focus on a key issue that to our knowledge has not been addressed—namely the strong existing counter-incentives for original authors to have their work replicated. We propose a solution for increasing the frequency of well-conducted and appropriate replications. In our view, the academy is lacking not only incentives for the replicator, but also for the authors of an initial, original, study to accommodate a possible replication. For instance, Ioannidis (2014) notes that sharing the materials necessary for allowing replications to take place is dis-incentivized in the current academic system. In empirical economics, Camfield and Palmer-Jones (2013) lament the lack of incentives to share tacit knowledge necessary for replication, and Andreoli-Versbach and Mueller-Langer (2013) document that economics researchers are loathe to sharing their data voluntarily.

Importantly, in this paper we advocate in favour of institutional reforms that ensure a significant expected benefit to original authors from a replication attempt. This might involve novel solutions that require an attitude change not only by authors, but also by influential stakeholders such as professional associations (e.g., the Economic Science Association) funding agencies (such as the European Research Council), academic departments, and the like. For instance, in order to achieve higher expected benefits for the original authors, scientific journals (especially those run under the auspices of professional associations) should always allow for timely commentaries after a replication attempt. This will require very little journal space while allowing some return in terms of publications and citations to the original authors.[18] Moreover, this practice would allow the careful examination of factors leading to the replication results, thus contributing to knowledge accumulation.[19]

In addition, we support the use of the number of replications of one's experimental work (which we might call "replicitations") as a metric for one's research quality. We find this appropriate since it is clear that if one's work is interesting, then it should attract replication efforts. Instead of counting citations alone, replicitations should also be counted and taken

[17] We can testify this from recent personal experience, being on *both* sides of the fence! (See Maniadis Tufano and List, 2014; Simonson Nelson and Simmons, 2014; and Maniadis Tufano and List, 2015).

[18] We expect that citations of commentaries along with original and replication articles should be appropriately enforced by authors, reviewers and editors.

[19] A nice example of this is illustrated in the commentaries that appeared in the psychology journal *Social Psychology* (volume 45, 2014, issue 4), which followed the (registered) replication reports published in the previous issue of that same journal.

into account for funding and promotion purposes, etc.[20] We further argue that 'adversarial collaborations' (in which two research teams proposing contrasting hypothesis attempt to resolve the dispute empirically – e.g., Bateman et al., 2005) or 'daisy chains' of replications (in which a given lab attempts to replicate a finding by its 'neighbour' lab while being supervised by a member of the replicated lab – e.g., Young, 2012) should be used more often to cast light on "true" empirical regularities.

Reforms along the lines outlined above will likely strike many birds with one stone: first, the practice of replication shall no longer make enemies, but friends. Second, the quality of replications will increase, since the original authors will devote more time and effort to help with the replication. Third, since, as we envision, a replication could be viewed as an act of goodwill, this might induce people to replicate more. In addition, we hypothesize that if there is a shift into a regime of widespread replication, people will no longer be stigmatized if one or two of their findings are not replicated. For us, noise is unavoidable in behavioural research, and more importantly there are also canonical and inevitable forces in scientific practices (outside the realm of fraud) that may tend to bias results (Nosek, Spies and Motyl, 2012).

Conclusions

In this paper we contribute to the lively discussion concerning the necessary reforms of scientific practices in the face of the 'research reproducibility problem'. We make three main points. First, increasing the incentives for replication is just one of the four types of reform proposals that researchers in several disciplines have made. Second, our state of knowledge regarding what these proposals entail could greatly benefit from employing the formal tools of economic analysis, and we illustrate this point using distinct examples. This means that economic theorists should play an active role, but they should also make a serious effort to collaborate in multidisciplinary teams and fuel a virtuous circle between theory and empirical evidence. Third, any solution proposed so far has under-appreciated a potential key variable that distorts scientific practice: the enmity among researchers that replication induces. We propose a set of novel solutions to this problem, noting that formal evaluation of these proposals is also necessary in order to fully appreciate the trade-offs involved.

REFERENCES

Andreoli Versbach, P., & Mueller-Langer F. (2013). Open access to data: An ideal professed but not practised. *RatSWD Working Papers, 215*. Retrieved March 2015.

Bateman, Ian, Kahneman, D., Munro, A., Starmer, C., & Sugden, R. (2005). Testing competing models of loss aversion: an adversarial collaboration. *Journal of Public Economics* 89(8), 1561–1580.

Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature* 483.7391, 531-533.

[20] This may be obtained for instance by introducing one or more JEL codes dedicated to replications studies, which will make the meta-database searches much easier and facilitate the use of relevant information by funding agencies, promotion panels, award committees and the like. An added benefit is that this will also facilitate careful meta-analytic research using the search engines.

Bettis, R. A. (2012). The search for asterisks: compromised statistical tests and flawed theories. *Strategic Management Journal* 33(1) 108-113.

Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.

Camfield, L., & Palmer-Jones, R. (2013). Three 'Rs' of Econometrics: Repetition, Reproduction and Replication. *Journal of Development Studies* 49(12), 1607-1614.

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton, NJ: Princenton University Press.

Coffman, L., & Niederle, M. (2014) Pre-Analysis Plans are not the Solution, Replications Might Be. Available at http://web.stanford.edu/~niederle/Coffman.Niederle.PAP.JEP.October2014.pdf

Dewald, W. G., Thursby, J.G., & Anderson, R. G. (1986). Replication in Empirical Economics: The Journal of Money, Credit and Banking Project. *American Economic Review* 76(4), 587-603.

Doshi, P., Goodman, S. N., & Ioannidis, J. P.A. (2013). "Raw data from clinical trials: within reach?" *Trends in Pharmacological Sciences* 34 (12), 645–647.

Dufwenberg, M., & Martinsson, P. (2014). Keeping Researchers Honest: The Case for Sealed-Envelope-Submissions." Mimeo.

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PloS One* 4(5), e5738.

Fanelli, D. (2013). Redefine misconduct as distorted reporting. *Nature* 494.7436, 149-149.

Gall, T., & Maniadis, Z. (2015). Evaluating solutions to the problem of false positives. University of Southampton Discussion Paper in Economics and Econometrics 1504.

Goldacre, B. (2011). Researchers don't mean to exaggerate, but lots of things can distort findings. *The Guardian*, August 2011.

Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences* 24 (03), 383-403.

Henry, E. (2009). Strategic disclosure of research results: the cost of proving your honesty. *The Economic Journal*, 119, 1036-1064

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine* 2.8, e124.

Ioannidis, J. P. A. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science* 7(6), 645-654.

Ioannidis, John PA. How to make more published research true. *PLoS Medicine* 11.10 (2014): e1001747.

Ioannidis, J. P. A., Allison, D. B., Ball, C. A., Coulibaly, I., Cui, X., Culhane, A.C., … & van Noort, V. (2009). Repeatability of published microarray gene expression analyses. *Nature Genetics* 41: 149-155.

Jennions, M. D., & Møller, A. P. (2002). Relationships fade with time: a meta-analysis of temporal trends in publication in ecology and evolution. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 269.1486, 43-48.

John, L. K., Loewenstein, G. & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological science* 23(5), 524-532.

Kahneman, D. (2014). A New Etiquette for Replication. *Social Psychology* 45(4), 299-311.

Kiri, B., Lacetera, N., & Zirulia, L. (2014). Above a Swamp: A Theory of High & Quality Scientific Production. Mimeo.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams Jr, R. B., Bahník, Š., Bernstein, M. J., ... & Woodzicka, J. A. (2014). Investigating variation in replicability: A "many labs" replication project. *Social Psychology*, *45*(3), 142.

Landis, S. C., Amara, S. G., Asadullah, K., Austin, C. P., Blumenstein, R., Bradley, E. W., ... & Silberberg, S. D. (2012). A call for transparent reporting to optimize the predictive value of preclinical research. *Nature*, *490*(7419), 187-191.

Lehrer, J.(2010). "The truth wears off." *The New Yorker* 13, 52.

List, J. A., Bailey, C. D., Euzent, P. J., & Martin, T. L. (2001). Academic economists behaving badly? A survey on three areas of unethical behaviour. *Economic Inquiry* 39(1), 162-170.

Maniadis, Z., Tufano, F., & List, J. A. (2014). One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects. *American. Economic Review* 104(1), 277-290.

Maniadis, Z., Tufano, F., & List, J. A. (2015). One Swallow Doesn't Make a Summer: Reply to Simonsohn, Nelson and Simmons (2014). Mimeo.

Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., ... & Van der Laan, M. (2014). Promoting transparency in social science research. *Science (New York, NY)*, *343*(6166), 30.

Moonesinghe, R., Khoury, M. J., & Janssens C. J. W. (2007). Most Published Research Findings Are False-But a Little Replication Goes a Long Way. *PLoS Medicine* 4(2): 0218-0221.

Nosek, B. A., & Lakens, D. (2014). Registered Reports. *Social Psychology* 45(3), 137-141.

Nosek, B. A., Spies, J. R. & Motyl, M. (2012). Scientific Utopia II. Restructuring Incentives and Practices to Promote Truth Over Publishability. *Perspectives on Psychological Science* 7(6), 615-631.

Poldrack, R. A., & Poline, J-B. (2014). The publication and reproducibility challenges of shared data. *Trends in Cognitive Sciences*.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* 22(11), 1359-1366.

Simonsohn, U., Simmons, J., &. Nelson, L. (2014). Anchoring is not a false-positive: Maniadis, Tufano, and List's (2014) 'failure-to-replicate' is actually entirely consistent with the original. Available at http://dx.doi.org/10.2139/ssrn.2351926.

The Economist. Trouble at the Lab. 19 October 2013: 23-27.

The Economist. Metaphysicians: Sloppy researchers beware. A new institute has you in its sights. March 2014.

Wacholder, S., Chanock, S., Garcia-Closas, M., El-ghormli, L., & Rothman, N. (2004). Assessing the Probability That a Positive Report is False: An Approach for Molecular Epidemiology Studies. *Journal of the National Cancer Institute* 96(6): 434-442.

Wagenmakers, Eric-Jan, & Forstmann, B. U. (2014). Rewarding high-power replication research. *Cortex* 51.10.

Young, E. (2012). Nobel laureate challenges psychologists to clean up their act. *Nature News* doi:10.1038/nature.2012.11535

Zimmer, C. (2011). It's science, but not necessarily right. *New York Times opinion. New York Times, New York, NY*.